

cui bono - computerunterstützte Inhaltsanalyse für die qualitative empirische Sozialforschung

Mohler, Peter Ph.

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Empfohlene Zitierung / Suggested Citation:

Mohler, P. P. (1992). cui bono - computerunterstützte Inhaltsanalyse für die qualitative empirische Sozialforschung. In J. H. P. Hoffmeyer-Zlotnik (Hrsg.), *Analyse verbaler Daten: über den Umgang mit qualitativen Daten* (S. 389-401). Opladen: Westdt. Verl. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-5695>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see: <https://creativecommons.org/licenses/by-nc-nd/4.0>

cui bono - Computerunterstützte Inhaltsanalyse für die qualitative empirische Sozialforschung

Peter Ph. Mohler

Cui, die computerunterstützte Inhaltsanalyse, ist der **algorithmisierte**, d.h. auf Computerregeln aufbauende, Teil der **sozialwissenschaftlichen** Inhaltsanalyse. Wie diese ist sie ein Verfahren zur systematischen Informationsreduktion von Textbedeutungen. Dazu bedient man sich in der cui vielfältiger, insbesondere quantitativ-statistischer Verfahren, nachdem man bestimmte Wörterkategorien eines Textes ausgezählt hat. Für qualitative Forschungsansätze bietet die cui neben der Möglichkeit, Hypothesen systematisch zu prüfen, Verfahren für die Textexploration an. Diese Explorationsverfahren werden in der cui für die Vorbereitung von Kategorienschemata benötigt. In der qualitativen Forschung können diese Verfahren aber ebenso gut für die Suche nach relevanten Textstellen oder die Übersicht über das Vokabular von Sprechern verwendet werden. Damit kann die cui auch zur Verknüpfung qualitativer und quantitativer Forschungsansätze beitragen.

1. Problemstellung und Übersicht

Die computerunterstützte Inhaltsanalyse (cui) ist ein Zweiggebiet der "Inhaltsanalyse" oder "Content Analyse" innerhalb des Paradigmas der quantitativen empirischen Sozialforschung. Das quantitative Paradigma war lange auf *Verteilung* von *Merkmalen* in *Populationen* gerichtet. Typologische Aussagen, wie sie im qualitativen Paradigma vorherrschen,

standen nicht im Vordergrund der Untersuchungen quantitativ vorgehender Forscher. Diese Differenzierung gehörte zu den tiefsten Kluft, die qualitative und quantitative empirische Sozialforschung trennten und sie wirkt noch fort, auch wenn quantitative via Strukturgleichungsmodellen u.ä. sich heute besonders Typologien zuwenden (vgl. Mohler 1981). Dennoch **muß** die alte Kluft hier, bei der Betrachtung des Nutzens der *cui* für qualitative Forschung, beachtet werden. Die *cui*-Verfahren wurden für verteilungsbezogene Aussagen entwickelt. Abgesehen von Nebeneffekten, die durch geschickte Umwidmung einzelner Verfahren für qualitative Ansprüche gewonnen werden können, dienen *cui*-Verfahren der formalen Prüfung qualitativ gewonnener Typologien (vgl. u. a. Mohler 1989). Und damit bilden sie die erkenntnistheoretisch wie methodologisch notwendige Brücke zwischen hermeneutischer und intersubjektiv nachprüfbarer Erkenntnis mit dem bekannten Gefälle an Erkenntnisgewinn.

Um den Nutzen der *cui* für qualitativ vorgehende Forscher besser bewerten zu können, wird deshalb nach einer kurzen Begriffsdefinition ein Überblick über die Hauptströmungen in der *cui* gegeben.

Unter dem Begriff "Inhaltsanalyse" werden unterschiedliche quantitative sozialwissenschaftliche Verfahren der Textanalyse zusammengefaßt. Ihnen allen sind zwei Merkmale gemeinsam: Erstens wird immer von einer Textsequenz auf Kontextmerkmale geschlossen (Autor, Leser und gesellschaftliches Umfeld der Textproduktion und **-rezeption** bilden die drei wesentlichen Untereinheiten der jeweiligen Kontextanalyse). Zweitens wird in der Inhaltsanalyse prinzipiell ein hoher Grad der Systematisierung angestrebt. Durch die Systematisierung soll die Intersubjektivität der Vorgehensweisen gesichert werden. Probleme, die mit einer rigiden Handhabung dieses Paradigmas auftreten, zum Beispiel trivialer Quantifizierungen oder unangemessener Formalismen, treten bei der Analyse von Texten verstärkt auf (Mochmann 1985). Wesentlich ist für die Inhaltsanalyse ein weiterer Aspekt, nämlich der der Informationsreduktion: Es ist in der Regel nicht das Anliegen quantitativer sozialwissenschaftlicher Forschung, soziale Fakten zu paraphrasieren; vielmehr sollen aus der Unzahl möglicher Merkmale diejenigen miteinander in Verbindung gesetzt werden, die wesentlich zur Erklärung bestimmter Phänomene beitragen und es sollen deren Verteilungen bestimmt werden. In der Inhaltsanalyse wurden nun Verfahren entwickelt, die eine intersubjektiv überprüfbare Reduktion von Textmerkmalen auf die für eine angemessene Kontexterklärung notwendige Zahl leisten.

Die Verfahren der konventionellen Inhaltsanalyse erfordern immer, daß Vercoder Teile des Textes bestimmten Merkmalen (Kategorien) zuordnen. In der Regel beruht eine Vercodung auf Transkripten oder originären Texten. In der computerunterstützten Inhaltsanalyse erfolgen die von

Vercodern zu leistenden Merkmalszuordnungen mit Hilfe von Computer-Programmen. Eine Verarbeitung von manuell erzeugten Merkmals-häufigkeiten o. ä. durch statistische Programme ist nach dieser Definition keine computerunterstützte Inhaltsanalyse im engeren Sinne; obgleich man sich künftig Anwendungen vorstellen kann, wo manuell identifizierte, komplexe Textmerkmale als Kategorien in Wissensbasen der Künstlichen Intelligenz verarbeitet werden. Computerunterstützte inhaltsanalytische Verfahren werden in den unterschiedlichsten sozialwissenschaftlichen Bereichen eingesetzt. So findet man Studien über Zeitungen (Deweese 1976, Mohler 1985), Gruppengespräche (Aries 1977, Grunert & Bader 1985), psycho-analytische Gespräche (Gottschalk 1985, Kächele & Mergenthaler 1984), historisch-politische Texte (Namenwirth & Weber 1987) und literarische Texte (Walker 1975, Martindale 1986). Sie haben sich sowohl für die thematische Erschließung als auch für die hypothesenzentrierte Analyse von Texten bewährt. Bestimmend für die weitere Entwicklung der computerunterstützten Inhaltsanalyse ist die Verbreitung und Verbesserung der bestehenden Programmpakete und die Ergänzung der verwendeten Algorithmen um syntaktische und kontextsensitive semantische Komponenten.

2. Ansätze der computerunterstützten Inhaltsanalyse

Innerhalb der computerunterstützten Inhaltsanalyse gibt es zwei prinzipiell unterschiedliche Ansätze. Einmal den sogenannten a-priori-Ansatz, der sich auf allgemeine Klassifikationsschemata stützt und zum anderen den empirischen Ansatz, der induktiv gewonnene Klassifikationsschemata einsetzt. Beide Ansätze führen im Prinzip kontextfreie Einwortvercodungen durch. - Der a-priori-Ansatz wurde von einer Arbeitsgruppe um Philip J. Stone Anfang der sechziger Jahre entwickelt (Stone et. al. 1966). Dieser Ansatz geht von der Idee der Übersetzung einer Objektsprache in eine Metasprache aus (Mochmann 1980: 13). Dabei wird ein sehr einfacher Begriff der Übersetzung verwendet, indem eine weitgehend kontextfreie Zuordnung von Textwörtern zu sozialwissenschaftlichen Begriffen als "Übersetzungsprozeß" verstanden wird. Die Zuordnung geschieht durch das Abgleichen eines inhaltsanalytischen Wörterbuchs (Diktionärs) mit dem Textkorpus. Das Wörterbuch besteht aus nach sozialwissenschaftlichen Begriffen (Kategorien) angeordneten Wortlisten, ähnlich einem Thesaurus oder der Auflistung semantischer Felder. Wird ein bestimmter Eintrag des Wörterbuchs in einer Texteinheit gefunden, so wird der Zähler für die betreffende Kategorie um eins erhöht. Ein einfaches Beispiel für eine

inhaltsanalytische Kategorie wäre folgende Liste von Wörtern: "demokratisch, Demokratie, Demokrat", die unter einer Kategorie Demokratie **zusammengefaßt** werden könnten. Es lassen sich aber auch komplexere Beispiele angeben, wie zum Beispiel eine Kategorisierung von Verben nach Aktionsarten oder Angsttypen. Ein inhaltsanalytisches Wörterbuch, das alle für eine bestimmte Analyse notwendigen Kategorien enthält, wird auch als Klassifikationsschema bezeichnet. - Grundsätzlich ist der Basisbegriff des a-priori-Ansatzes, die Übersetzung von einer Objekt- in eine Metasprache, zu kritisieren. Denn eigentlich beinhaltet der Begriff der Übersetzung eine Reziprozität im Sinne der immer gegebenen Möglichkeit einer Rückübersetzung. Die Reduktion der Lexeme eines Textes auf wenige **sozialwissenschaftliche** Begriffe, erlaubt dagegen keine Rekonstruktion des Textes. Hier wäre es wohl besser, von Abbildungen der Textwörter auf ein **Klassifikationsschema** als von einer Übersetzung zu sprechen. - Der empirische Ansatz wurde insbesondere von Howard P. Iker entwickelt (Iker Harway 1969). In diesem Ansatz, der ebenfalls kontextfrei Wörter klassifiziert, werden empirisch gegebene Assoziationsmuster von Wörtern als Kategorien interpretiert. Im Prinzip bildet hier jedes einzelne Wort eine eigene Kategorie; in der Praxis werden jedoch Synonym- und **Lemmatisierungslisten** eingesetzt, um die Häufigkeiten einzelner "Worte" zu erhöhen. Für alle so definierten Wörter wird pro Texteinheit ein bestimmter **Assoziationskoeffizient** (i.d.R. Pearsons r , Eta) berechnet, der die Häufigkeit des gemeinsamen **Auftretens zweier** Wörter angibt. Mit Hilfe multivariater statistischer Analyseverfahren, wie zum Beispiel Cluster- und **Faktorenanalysen** oder multidimensionaler Skalierungen **können** dann **Kategorienmuster** auf der Basis der Koeffizientenmatrix induktiv gewonnen werden. Diese Kategorienmuster beschreiben Ereignisse oder Themen in einem Korpus. Auch wenn die **Synonymlisten** und später die Kategorienmuster auf den ersten Blick den Kategorien des Stoneschen a-priori-Ansatzes ähnlich sind, so ist doch festzuhalten, daß letztere theoretische Konzepte bezeichnen, die für mehr als einen Korpus Gültigkeit beanspruchen (Mochmann 1985, Mohler 1985). Kritisch ist hier zu bemerken, daß der empirische Ansatz nur für Textkorpora geeignet ist, bei denen alle Texteinheiten zumindest in sich hochredundant bezüglich der Wortwahl und der Thematik sind. - Neben diesen **beiden** genannten Ansätzen gibt es Mischformen oder Varianten, die aber prinzipiell auf den a-priori- oder den empirischen Ansatz **zurückführbar** sind.

3. Verfahrenstechniken

Eines der wesentlichen Probleme in der Inhaltsanalyse im allgemeinen und der computerunterstützten im besonderen sind die Schlüsse, die vom Text auf einen Kontext gezogen werden. Wie unter Punkt 1 erwähnt, sind Autor, Leser und gesellschaftliches Umfeld der Textproduktion und **-rezeption** die drei wesentlichen Einheiten von Kontextanalysen. Deshalb ist in der Inhaltsanalyse neben dem immanenten Textverständnis auch das jeweilige **Weltwissen** für die Ergebnisse einer Untersuchung bestimmend. Im Falle der Inferenz von einem Text auf den Autor befindet man sich noch auf einem relativ sicheren Grund; dieser Grund wird immer unsicherer, je weiter man sich räumlich und zeitlich von dem Urheber entfernt (Hockey 1980: 122 ff). Für Inferenzen von Texten auf die Leserschaft oder ein bestimmtes gesellschaftliches Umfeld sind erhebliche theoretische und empirische Probleme zu überwinden, weil der Blick auf die Leserschaft und auf das gesellschaftliche Umfeld immer doppelt, sowohl durch den Text als auch durch dessen Autor, gebrochen wird (Weber 1985: 18 ff, 58 ff). Die sich hieraus ergebenden Probleme der Gültigkeit inhaltsanalytischer Aussagen sind noch nicht abschließend geklärt.

3.1. Textauswahl

Das Problem der Textauswahl ist direkt mit dem der Inferenzen verbunden. Mit Ausnahme der seltenen Fälle, in denen alle für eine Untersuchung relevanten Texte erhoben werden können (Totalerhebung), müssen für **sozialwissenschaftliche** Inhaltsanalysen untersuchungsspezifische, repräsentative Textkorpora zusammengestellt werden. Hier ist **dann** immer die Frage: "repräsentativ in Bezug auf was?" zu stellen. Dieses Problem tritt hier in ähnlicher Weise auf, wie in der linguistischen Textdeskription, wo nicht immer eindeutig ist, welcher Teil der Sprache durch einen bestimmten Korpus repräsentiert wird (Devons 1985). Oftmals tritt das Problem nur verdeckt auf, weil scheinbar Totalerhebungen vorgenommen wurden. Zum Beispiel ist die Analyse aller Reden des britischen Monarchen zur Parlamentseröffnung seit dem 17. Jahrhundert unter Umständen wieder nur eine Auswahl, nämlich aus vielen anderen denkbaren Texten, die auf das politische System verweisen (Namenwirth & Weber 1987). - Auch rein zufällige Auswahlverfahren garantieren keineswegs eine spätere Repräsentativität der Stichprobe für bestimmte Zwecke. So ist zum Beispiel das in der Kommunikationswissenschaft öfters angewandte Verfahren der "künstlichen Woche" (bei dem aus einer ersten Woche der Montag als

Stichtag, aus einer zweiten Woche der Dienstag und schließlich aus einer siebten Woche der Sonntag als Stichtag genommen wird) für eine vergleichende Inhaltsanalyse von Tageszeitungen nicht unbedingt angemessen. Dies liegt **daran**, daß bestimmte Nachrichten in überregionalen Blättern früher als in regionalen Blättern und umgekehrt **aufreten**. Würde die Stichprobe mit Hilfe des Verfahrens der "künstlichen Woche" gewonnen, dann wäre diese wegen dieser verzögerten Berichterstattung systematisch verzerrt und unter Umständen nicht mehr für einen gegebenen Untersuchungszweck repräsentativ. In jedem Falle muß also eine genaue Prüfung der Repräsentativität der auszuwählenden Texte einer Inhaltsanalyse vorausgehen. Rein mechanische Verfahren der Textauswahl sind immer problematisch, sie könnten allzuleicht zur Unbrauchbarkeit der Textauswahl führen.

3.2. Textdeskription und Textvercodung

Neben den in der Computerlinguistik entwickelten Deskriptoren, angefangen bei der Silbenlänge bis hin zur Satzstruktur (Hockey 1980), werden in der computerunterstützten Inhaltsanalyse Deskriptoren vornehmlich zur Exploration verwendet. Dazu gehören Worthäufigkeiten, Key-Word-In-Context-(KWIC-)Listen, Vokabularvergleichslisten und Konkordanzen (Weber 1985). In der Regel werden Worthäufigkeiten und KWIC-Listen für die Entwicklung und Validierung inhaltsanalytischer Wörterbücher oder von Synonymlisten eingesetzt. In seltenen Fällen dienen solche Deskriptoren auch analytischen Zwecken (Weber 1985, 44 ff.). - Die Textvercodung, das heißt die Zuordnung von Textelementen zu Kategorien beim a-priori-Verfahren oder **Synonymlisten** beim empirischen Verfahren, ist derzeit noch mit folgenden Problemen behaftet: In der Regel werden einzelne Wörter (auch Wortsequenzen) kontextfrei, **d.h.** ohne Rücksicht auf ihre syntaktische Funktion oder das Wortumfeld in dem sie stehen vercodet (vgl. Punkt 2). Damit stellt sich sofort das Problem der Homographen (**d.h.** gleiche Schreibweise, aber unterschiedliche Bedeutung durch Betonungsverschiebung bei der Aussprache: Tenor = Sänger, oder: Tenor = Leitsatz eines Gerichtsurteils) und Homonyme (gleiche Schreibweise, aber unterschiedliche Herkunft und andere Bedeutung: Steuer = Abgaben an den Staat, oder: Steuer = Lenkrad). Für das Englische liegt ein Lösungsvorschlag vor, der in der Praxis zu sehr befriedigenden Ergebnissen **führt** (Stone & Kelly 1976, Züll, Weber & Mohler 1989). Dieser Ansatz zielt nicht auf die Lösung aller Homographen und Homonymprobleme, sondern auf eine **befriedigende** Behandlung der häufigsten Wörter. Das von Stone und Kelly entwickelte und vom Benutzer erweiterbare System ordnet einzelne mehrdeutige

Wörter zu mehr als 90% der jeweils als zutreffend bezeichneten Bedeutung zu (Stone & Kelly 1976: 42). Zugleich handhabt dieses für den GENERAL INQUIRER (Stone 1966, Züll, Weber & Mohler 1989) entwickelte System auch die Abtrennung von Suffixen, d.h. von Pluralendungen oder bestimmten Endungen von Verben.

Beides, das Homographen/Homonyme- und das Suffixproblem, sind bis jetzt für andere Sprachen, insbesondere das Deutsche in einer für inhaltsanalytische Bedürfnisse befriedigenden Weise noch nicht gelöst (Mergenthaler 1985). Einerseits hängt dies mit spezifischen Strukturen der deutschen Sprache zusammen, die einen gegenüber dem Englischen wesentlich höheren Programmieraufwand erfordern, zum anderen mit den besonderen Bedürfnissen der Inhaltsanalyse, die ein weites Spektrum an wissenschaftlichen Interessen und damit an Textsorten abdecken muß (vgl. Punkt 1). Befriedigende, vollautomatisierte Lösungen bestehen aber zur Zeit nur für spezifische Korpora, und der Aufwand für die Anpassung an andere Korpora ist noch so aufwendig, daß es oft einfacher ist, Homonymie Homographen per Hand, mit Hilfe von KWIC-Listen, aufzulösen und Suffixe mittels einfacher Wortstammdeklarationen auf grobe Weise abzutrennen.

Für die Überprüfung der maschinell durchgeführten Vercodung stehen im GENERAL INQUIRER und in TEXTPACK umfangreiche Prüfroutinen zur Verfügung. So ist es zum Beispiel auf sehr einfache Weise möglich, vergebene Codes in den Text zurückzuschreiben, die Verteilung von Kategorieneinträgen in Texteinheiten aufgrund von Vercodungen auszuwählen usw. Solche Routinen sind wegen der zuvor erwähnten großen Variabilität der analysierten Textkorpora notwendig; im Prinzip muß jeder Benutzer die für die Vercodung wesentlichen Teile der Programme manipulieren können, um so zu optimalen Ergebnissen für seinen spezifischen Untersuchungsansatz zu gelangen. Dazu gehört, daß er das Wörterbuch und das Korpus mittels Editoren frei manipulieren kann, aber auch, daß er gegebenenfalls zusätzliche Optionen selbst programmiert.

Ein weiterer Aspekt der Vercodung ist die nachfolgende Verarbeitung numerischer Codes mittels statistischer Programme. Bislang werden in der computerunterstützten Inhaltsanalyse zwei Formen numerischer Codes verwendet: einmal die Häufigkeit bestimmter Kategorien bezogen auf eine bestimmte Texteinheit, zum anderen die Abfolge von Kategorien innerhalb einer bestimmten Texteinheit. Vorherrschend werden **Kategorienhäufigkeiten** analysiert, wobei hier das Problem der Normierung der einzelnen Häufigkeiten gelöst werden muß. Denn je länger eine Texteinheit ist, desto wahrscheinlicher ist, daß eine Kategorie mehrmals auftritt oder daß mehrere Kategorien in ihr gemeinsam auftreten. Wenn Texteinheiten miteinander verglichen werden sollen, dann muß die Verzerrung, die aus

der Textlänge entstehen kann, berücksichtigt, womöglich eliminiert werden. Dafür werden entweder die Texte in gleichlange Abschnitte eingeteilt oder - wo dies nicht sinnvoll ist, wie zum Beispiel bei Interviews mit Sprecherwechsel - die **Rohwerte** der Kategorienhäufigkeiten in Prozentwerte umgerechnet. Problematisch ist bei letzterem Vorgehen die jeweilige Wahl der Basis der Prozentuierung. Dies kann einmal die Summe aller in einem Abschnitt auftretenden Kategorien oder die Summe aller Wörter einer Kategorie sein. Für die zweite Art der Vercodung, die Sequenz von Kategorien, entsteht ein besonderes Problem, wenn man Abstände zwischen einzelnen Kategorien bestimmen will. Es ist dann nämlich zu fragen, ob die Abstände sich alleine auf die vergebenen Kategorien beziehen sollen, oder ob auch die tatsächliche Zahl von Wörtern, die zwischen solchen Wörtern liegen, die Kategorien zugeordnet werden können, berücksichtigt werden sollen. Für dieses, wie für das Problem der Normierung bei Häufigkeitsauszählungen, liegen keine verbindlichen, theoretisch begründeten Empfehlungen vor. Ähnlich wie bei der **Textauswahl** und der Wahl der multivariaten Analyseverfahren, auf die hier nicht eingegangen werden soll (vgl. Weber 1985: 58 ff), müssen jeweils untersuchungsspezifische Lösungen angestrebt werden.

4. Verknüpfung von computerunterstützter Inhaltsanalyse und qualitativen Textanalysen

Aus den oben skizzierten Ansätzen ergibt sich ein weiterer Bereich, in dem Verfahren der computerunterstützten Inhaltsanalyse in der qualitativen Forschung Anwendung finden können.

Ein erster großer Bereich ist der der Texterschließung. Typischerweise werden in qualitativen Forschungsansätzen schwach strukturierte Datenmengen, insbesondere Texte, generiert. Bestimmte cui-Verfahren können dann benutzt werden, um die Daten systematisch zu erschließen. Dies beginnt mit einfachen Suchprozessen, bei denen nach bestimmten Wörtern (z. B. Namen) im Text gesucht wird. Während faktisch fast jedes Textverarbeitungsprogramm (neudeutsch: "Wordprocessor") nach bestimmten Wörtern suchen kann, kann fast keines automatisch jede Fundstelle auch ausdrucken. Dazu benötigt man die in der cui- und der Computerlinguistik üblichen KWIC-(Key-Word-In-Context) Programme. Ebenso können cui-Programme benutzt werden, um große Konkordanzen anzulegen.

Mit cui-Computerprogrammen wie z. B. **TEXTPACK** kann man weiterhin externe Informationen benutzen, um Texte zu selektieren und um sie neu

zusammenzustellen. Beide Vorgehensweisen können miteinander kombiniert werden, wodurch man die Möglichkeiten, einen Text nach verschiedensten Kriterien zu erschließen, vervielfachen kann. Dazu ein Beispiel: Nehmen wir an, wir hätten fünf Gruppendiskussionen zum Thema "Stadt-sanierung" durchgeführt. Wir wissen, daß die "Schicki-Micki-Sanierung" das Hauptthema der Diskussionen war. Nebenbei wurde auch über die Kindergartensituation diskutiert. Mit Hilfe von KWIC finden wir schnell alle Textstellen in denen von "Kindergarten", "Hort", "Tagesstätte" usw. gesprochen wurde. Nun haben wir 100 verschiedene Fundstellen. Um die Analyse zu systematisieren, wollen wir die Äußerungen von Frauen mit kleinen Kindern gegenüber Frauen ohne kleine Kinder kontrastieren, sowie noch nach Viel- und Wenigsprechern unterscheiden. Um letzteres herauszufinden zählen wir einfach alle Wörter, die jede Person gesprochen hat und legen ein Kriterium fest, das Viel- von Wenigsprechern trennt, und fügen das Ergebnis zu den Daten, die wir über jeden Sprecher besitzen (Alter, Geschlecht, kleine Kinder, usw.). Damit haben wir vier Kontrastgruppen definiert (vgl. Tab. 1). Dann lassen wir von dem Programm vier neue Texte, d. h. je einen pro Gruppe, zusammenstellen und schauen uns dann im einzelnen die entsprechenden Textstellen an.

Tabelle 1: *Beispiel: Auswahl von vier Kontrastgruppen*

Vielsprecher		Wenigsprecher	
!-----!			
Frauen mit kl. Kinder	Frauen ohne kl. Kinder	Frauen mit M. Kinder	Frauen ohne kl. Kinder

Im Prinzip haben wir nichts anderes getan, als Textstellen herauszuziehen, auf Karteikarten zu kleben und diese Karten bestimmten Gruppen zuzuordnen. Statt mit Kopierer, Schere, Kleber und Karteikarten mit dem Computer zu hantieren lohnt sich natürlich nur dann, wenn wir viele solcher Textpermutationen machen wollen.

Ein zweiter großer, in der qualitativen Forschung bis jetzt faktisch völlig mißachteter Bereich, ist die Nutzung von cui-Verfahren zur formalen Prüfung von hermeneutisch-interpretativ gewonnenen Ergebnissen. Man kann nämlich sehr genau beschreiben, wie ein bestimmter Typ sich verhalten sollte oder welche Textmerkmale einen Typus konstituieren. Auch dazu ein Beispiel: Eine Interpretation von Texten der Frankfurter

Allgemeinen Zeitung ergab, daß man zwischen 1949 und 1984 zwei Perioden unterschiedlicher Thematisierungen von Werten annehmen kam. Um diese interpretativ gewonnene Unterscheidung formal zu überprüfen, wurde zuerst jeder Text verschlagwortet, d. h., es wurde notiert, welche Werte in ihm thematisiert wurden, und dann wurde mit Hilfe eines statistischen Verfahrens versucht, die Texte der einen oder anderen Periode *alleine an Hand der Kombination von Schlagwörtern* zuzuordnen. Dies gelang in über 90% der Fälle (Mohler, 1989).

Mit solchen formalen Absicherungen können hermeneutisch-interpretative Ergebnisse - auch Typologien - gegen den allgegenwärtigen Vorwurf der Beliebigkeit, den Quantitative gerne vorbringen, abgesichert werden. Nebenbei bemerkt, die Ergebnisse qualitativer Sozialforschung lassen sich oft stringenter formalisieren als die pseudoformalen Häufigkeitsverteilungen der Quantiativen. Hier ist also noch ein großer Bereich, wo Verfahren der cui nützlich (*bono*) für die qualitativ orientierte Sozialforschung sein können.

5. Zusammenfassung

Die cui, als Verfahren für die Untersuchung großer, relativ stark strukturierter Textmengen eignet sich für eher qualitativ orientierte Forschungsprojekte zum einen als Instrument der thematischen Exploration. Dazu könnten die o.a. empirischen Verfahren a la Iker benutzt werden. Weiterhin ist sie ein gutes Hilfsmittel für die systematische Überprüfung qualitativ gewonnener Hypothesen über einen Text, denn auch metatheoretische Regelmäßigkeiten bedürfen der sprachlichen Objektivierung durch Wörter und deren spezifischer Muster. Für die Hypothesenprüfung oder auch die systematische Suche nach Textbelegen eignen sich die o.a. a-priori-Verfahren.

Es sei hier aber festgehalten, daß die computerunterstützte Inhaltsanalyse als das vollalgorithmisierte Verfahren der klassischen sozialwissenschaftlichen Inhaltsanalyse vom Forscher einerseits genaue Vorstellungen von dem, was untersucht werden soll, erfordert; andererseits erlaubt ihre strenge Systematik auch präzisere Aussagen über das, was in einem Text offensichtlich sein soll und was nicht. Damit kann die cui auch für qualitativ orientierte Forschung zusätzlichen Erkenntnisgewinn bringen.

Literatur

- Aries, E. F., 1977, Male-female interpersonal styles in all male, female, and mixed groups; in: Sargent, A. G. (Hg.), 1977
- Baumann U. (Hg.), 1984, Psychotherapie: Makro-/Mikroperspektive. Göttingen
- Devons, N., 1985, Concerning Corpus Derived Lexical Frequency Ratings; in: Preprints for the XIIth ALLC Conference, Juni 1985. Nizza
- Deweese, L. C., 1976, A Computer Content Analysis of printed Media: A Limited Feasibility Study; in: Public Opinion Quarterly 40:1976, S. 92-100.
- Gerbner, G., u. a. (Hgg.), 1969: The Analysis of Communication Content. New York
- Gottschalk, L. A., 1985, A Note on Computer Scoring of Verbal Content Analysis; in: Sprache und Datenverarbeitung 9/2/1985, S. 29-35
- Grunert, K.G. & Bader, M., 1985, Die Weiterverarbeitung qualitativer Daten durch computerunterstützte Inhaltsanalyse; in: Marketing 4/1985, S. 238-247.
- Hockey, S., 1980, A Guide to Computer Applications in Humanities. Baltimore
- Iker, H. J. & N. I. Harway, 1969, A Computer Systems Approach Toward the Recognition and Analysis of Content; in: Gerbner, G., u. a. (Hgg.), 1969, S. 381-406
- Kächele, H. & E. Mergenthaler, 1984, Auf dem Wege zur computerunterstützten Textanalyse in psychotherapeutischen Prozessforschung; in: Baumann U. (Hg.), 1984, S. 223-229.
- Langner R. (Hg.), 1986, Psychologie der Literatur. Theorien, Methoden, Ergebnisse. Weinheim

- Martindale, V., 1986, Psychologie der Literaturgeschichte; in: Langner R. (Hg.), 1986
- Mergenthaler, E., 1985, Anwendung der Textanalyse bei gesprochener Sprache am Beispiel psychoanalytischer Gespräche; in: Sprache und Datenverarbeitung 9/2/1985, S. 16-20
- Mochmann, E., 1985, Inhaltsanalyse in den Sozialwissenschaften; in: Sprache und Datenverarbeitung, 9/2/1985, S. 5-10
- Mochmann, E. (Hg.), 1980, Computerstrategien für die Kommunikationsanalyse. Frankfurt a. M.
- Mohler, P. Ph., 1981, Zur Pragmatik qualitativer und quantitativer Sozialforschung; in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 33/1981, S. 716-734
- Mohler, P. Ph., 1985, Computerunterstützte Inhaltsanalyse - zwischen Algorithmen und Mythen; in: Sprache und Datenverarbeitung, 9/2/1985, S. 11-15
- Mohler, P. Ph., 1989, Wertkonflikt oder Wertdiffusion: Ein Vergleich von Umfragen in der Bevölkerung und der Inhaltsanalyse von Leitartikeln der FAZ; in: Kölner Zeitschrift für Soziologie und Sozialpsychologie 41/1989, S. 95-122
- Namenwirth, J. Z. & R. P. Weber, 1987, *Dynamics of Culture*. Boston
- Sargent, A. G. (Hg.), 1977, *Beyond Sex Roles*. St. Paul, MN
- Stone, Ph. J. et. al., 1966, *The General Inquirer - A Computer Approach to Content Analysis*. Cambridge, Mass.
- Stone, Ph. J. & E. Kelly, 1976, *Computer Aided Recognition of English Word Senses*. Amsterdam
- Walker, A.W., 1975, *The Empirical Delineation of Two Musical Taste Cultures: A Content Analysis of Best-Selling Soul and Popular Recordings from 1962-1973*, New School Social Research, MS-vervielfältigt (Ph.D. Diss.)

Weber, R. P., 1985, Basic Content Analysis. Sage Series: Quantitative Applications in the Social Sciences 49. Beverly Hills

Züll, C., R. P. Weber & P. Ph. Mohler, 1989, Computer Aided Textclassification for the Social Science: The General Inquirer III. Mannheim